Predicting the Religion of a Country from the colors in its Flag using Data Mining Techniques

Rama Krishna Kallam
Computer Science Dept.
University of Bridgeport, Bridgeport, CT

Abstract

Huge amounts of data is being transferred ever since the evolution of internet and social media. By observing that data in an efficient way we can predict a lot of things early based on the past data. Many data mining techniques and algorithms are available to obtain results from data. In this project I aim to predict the religion of a country based on the colors in its flag. National flags are designed according to the culture and traditions of the people. The colors and symbols in the flags represent those. Classification and clustering algorithms are used and compared. The predictions may or may not become true. But the algorithm that has a high percentage of correct predictions is considered the best.

Problem Statement

The religion of the country should be predicted based on the predominant colors in its National flag. Each flag has at least two different colors in it. We consider the color which occupies most of the flag. Classification and clustering techniques are suitable for processing this kind of data. Classification is for supervised learning i.e., the training data is accompanied by labels indicating the class of observations. Clustering is for unsupervised learning i.e., the data will have more intra-cluster similarity and less inter-cluster similarity. We consider Naïve Bayes, K-Star, PBF Network and J48 algorithms for classification and k-means algorithm for clustering the data.

Experiments

The flags dataset considered is having the data of 194 countries each as an instance in prediction. The other attributes include name, landmass, zone, area, population, language, bars, stripes, colors, religion, sunstars, red, green, blue, gold, white, black, orange, mainhue, circle, crosses, saltires, crescent, triangle, icon, animate, text, top-left and bottom-right.

Classification:

The flags data is processed to Naïve Bayes, K-Star, PBF Network and J48 algorithms and the results are noted.

Total instances = 194

Algorithm	Naïve Bayes	K-Star	PBF Network	J48
Correctly classified instances	150	194	190	160
Incorrectly classified instances	44	0	4	34
Relative absolute error	30%	0%	6.2785%	26.6751%
Root relative squared error	66%	0.0002%	24.7055%	51.7406%

Clustering:

Using K-means algorithm for clustering with 5 clusters. Initially 5 data points are assumed as centers for their respective clusters. The points are allotted to the cluster whose center is nearer to the point. After the first round of allotment the mean of each cluster will become new center for each cluster. Again the data points are allotted to the cluster whose new center is closest. After this second round of allotment the five clusters may or may not have the same data points as they had after the first allotment. If the clusters have same data then that is the final clustering. Otherwise iterations have to be done in the same way as mentioned above until two consecutive allotments have same data in their respective clusters.

Results

For classification:
From the table it is clearly seen that K-Star outperformed J48, Naïve bayes, PBF Network algorithms. It predicted all the instances correctly.

For clustering:
1st cluster: 51 countries. Belong to North America, speak English, religion is other Christian, uses red, blue, white colors in flag, mainhue color is blue.
2nd cluster: 37 countries. Belong to Africa, speak other language, religion is ethnic, use red, green, gold and mainhue color is red.
3rd cluster: 31 countries. Belong to Asia, speak Arabic, religion is Muslim, uses red, green white colors and mainhue color is green.
4th cluster: 41 countries. Belong to Europe, speak English, religion is catholic, uses red, blue, white colors in flag and mainhue color is red.
5th cluster: 34 countries. Belong to Asia, speak other languages, religion is muslim and marxist,uses red and white colors in flag and mainhue color is red.

Conclusion

We can predict the language of the country as well just the way as the religion is predicted. In future we can predict the change in economic status of a country based on the previous years' financial data. The population growth of the country can also be predicted using any of the data mining techniques.

Acknowledgements

The studies were supported by Prof. Jeyongkyu Lee, Computer Science Engineering Dept., University of Bridgeport